

Lei XIA

xialeibrian@outlook.com | 86 13866650918 | 852 64702390

NAN HAI MANSION, Kennedy Town, Central and Western District, Hong Kong

EDUCATION BACKGROUND

Hong Kong Baptist University

Hong Kong, China

PhD in Computer Science

09/2024-08/2028

- **Research Interests:** LLM, Social Computing, Misinformation.
- **Supervisor:** Dr. Yupeng Li (HKBU IDM), Dr. Hongning Dai (HKBU CS), Prof. Francis C. M. Lau (HKU)

The University of Hong Kong

Hong Kong, China

MSc in Computer Science

09/2022-07/2024

- **GPA:** 3.65/4.30 (Top 5%, Distinction)
- **Core Modules:** Natural Language Processing, Artificial Intelligence of Things, Data Mining, Visualization and Visual Analytics, Deep Learning, Topic in Artificial Intelligence, Multimedia Technologies, etc.
- **Dissertation Title:** Using Pre-trained Language Model for Accurate ESG Prediction.
- **Supervisor:** Dr. Qi Liu

Hong Kong Baptist University

Zhuhai, China

Bachelor of Science (Honours) in Computer Science and Technology

09/2018-06/2022

- **GPA:** 3.53/4 (Top 10%, First Class); 3.75/4 (Major)
- **Core Modules:** Object-Oriented Programming, Data Structures and Algorithms, Computer Graphics, Digital Image Processing, Introduction to Machine Learning, Neural Networks and Deep Learning, Data Mining, etc.
- **Honors and Awards:** First Class Award for the 2021-2022 academic year (Intramural); First Class Award for the 2020-2021 academic year (Intramural); Second Class Award for the 2019-2020 academic year (Intramural); Certificate of Internship Scholarship (Intramural).

PUBLICATION

Using Pre-trained Language Model for Accurate ESG Prediction

Lei Xia, Mingming Yang, Qi Liu

In Proceedings of the Eighth Financial Technology and Natural Language Processing and the 1st Agent AI for Scenario Planning (pp. 1-22) (2024).

BiverWordle: Visualizing Stock Market Sentiment with Financial Text Data and Trends

Lei Xia, Yiping Gao, Le Lin, Yuxi Chen, Kang Zhang

The 16th International Symposium on Visual Information Communication and Interaction (VINCI 2023)

CONFERENCE PRESENTATIONS

Xia Lei and Cheng Haoyang. (2023). *Navigating Copyright and Data Transparency in AI-Generated Works: Legal and Technological Perspectives*. International Online Conference On Criminal Justice, Forensics And Victimization, India. March 25 - 26.

PROJECT EXPERIENCE

Document level Causal Relation Extraction

05/2023-02/2024

Supervised by Dr. Xinya DU (assistant professor at UTD.)

- Formulate CRE as a binary QA task with single-turn and multi-turn strategies.
- Leverage event structures constructed with relevant document-level IE models.
- Proposed a benchmark for multi-hop event-centric question and answering.
- Two papers are under review.

CUHK's Project on Pretrain Model - Temporal Pre-training Language Model

07/2022-01/2023

Supervised by Prof. Benyou Wang

- Improved the temporal sensitivity of the BERT model and explored the contextual properties of words under different temporal domains.
- Improved the sensitivity of year recognition by using n-bit encoding to represent the year. Took the smallest year as the base and set it to 0 and normalize all other years.
- Added a prompt for a domain, selectively masked more content related to time and domain, modified BERT's task to time-consistent loss and domain-consistent loss, and added time and domain token to CLS.

U.S. Patent Phrase to Phrase Matching (Kaggle Competition)

03/2022-06/2022

- Extracted relevant information by matching key phrases in patent files to determine the semantic similarity

between phrases.

- Adopted Bert For Patent + DeBERTa + ELECTRA + Funnel-Transformer for multi-model fusion, and added a linear layer and sigmoid.
- Received a **silver medal**, ranked 18th out of 1889 (**top1%**).

Financial Information Processing and Visualization (Undergraduate Dissertation) 05/2021-01/2022

Supervised by Prof. Kang Zhang (Research field: Visual Languages, Generative Art.); Final Grade: A.

- Visualize the financial information. Process the comment information of netizens in the financial forums, extract and visualize its emotional information.
- Perform financial textual data visualization based on AntV G2, built a visualization system with Django framework, and used themeRiver and world cloud to visualize financial themes over a period of time.
- Proposed a visualization theory called BiverWordle to visualize stock market sentiment with financial text data and trends.

INTERNSHIP EXPERIENCE

Tencent

Shenzhen, China

AI Lab - Natural Language Algorithms Group Platform Staff

07/2023-present

- Terminology Crawler Project for Term Translation Replacement Distributed Crawler Implemented with Scrapy and Redis.
- Completed four patents related to large language models.
- Develop a SFT dataset to improve Hunyuan LLM's translation ability by using terminology dictionary.
- Develop a neural network-based word alignment API, widely used in machine translation within the company.

PwC

Shenzhen, China

Tax Technology Intern

04/2022-06/2022

- Developed tax system interface for AAC Technologies Holdings Inc. Used .NET Interface Development and EF to realise the add, delete, change, and check functions of the database. Provided user interaction function using angularjs front-end framework.
- Got familiar with Sql server, operated and maintained the tax system, responsible for data query and data modification. Learned the development of Springboot and the application of Redis.

iFLYTEK Co., Ltd.

Hefei, China

Pre-sales Consulting Specialist of Intelligent Service Solutions Department

08/2020-09/2020

- Conducted research on financial technology in the industry, analyzing human-computer interaction, voice recognition, face recognition, video interaction, and predicting company's future in this field.
- Analyzed financial models of BAT and JD.com, and ecosystems such as Du Xiaoman and WeChat Pay.
- Managed operation and maintenance of department's document management website.

OTHER TRAINING

Duke Kunshan University

Kunshan, China

Course: Introduction to Machine Learning

08/2020

- Learned machine learning algorithms. Got full marks and a recommendation letter from Prof. Xin Li (Duke University).

Online Courses on Coursera

- **Machine Learning**, certificate issued by Stanford University
- **Deep Learning Specialization**, certificated issued by DeepLearning.AI
- **TensorFlow: Advanced Techniques Specialization**, certificated issued by DeepLearning.AI

SKILLS AND INTERESTS

- **Programming Language:** Python, Java, C.
- **Skills:** Natural Language Processing, Pre-training Language Model, Deep Learning.
- **IDE:** VS code, IDEA, Eclipse.
- **Deep Learning Frameworks:** TensorFlow, PyTorch.
- **Language:** Fluent in English, English Medium of Instruction.